



The
Patent
Office

097720588

PCT/GB 00/01614

INVESTOR IN PEOPLE

GBOO/01614

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

2/11

ESU

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 18 AUG 2000

WIPO

PCT

I, the undersigned, being an officer duly authorised in accordance with ~~Section 74(1) and (4)~~ of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

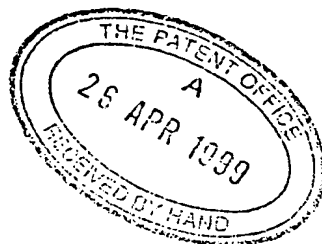
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 28 JUL 2000

BLANK PAGE

ff use only



27APR99 E442900-1 003274
P01/7700 0.00 - 9909605.9

26 APR 1999

Your reference Encoder Layering (B)

9909605.9

The
**Patent
Office**

Request for grant of a
Patent

Form 1/77

Patents Act 1977

1 Title of invention

Networked delivery of media files to clients

2. Applicant's details



First or only applicant

2a

If applying as a corporate body: Corporate Name

Telemedia Systems Limited

Country

Great Britain

2b

If applying as an individual or partnership

Surname

Forenames

2c

Address

Mount Pleasant House
2 Mount Pleasant
Huntingdon Road
Cambridge

UK Postcode

CB3 0RN

7647 860001

Country

Great Britain

ADP Number

☐

Second applicant (if any)

2d

Corporate Name

Country

2e

Surname

Forenames

2f

Address

UK Postcode

Country

ADP Number

3 Address for service

Agent's Name

Origin Limited

Agent's Address

24 Kings Avenue
London

Agent's postcode

N10 1 PB

7270457001

Agent's ADP
Number

C03274

4 Reference Number

Encoder layering (B)

5 Claiming an earlier application date

An earlier filing date is claimed:

Yes ☐

No ☒

Number of earlier
application or patent number

Filing date

15 (4) (Divisional)

8(3)

12(6)

37(4)

☐☐☐☐

6 Declaration of priority

Country of filing

Priority Application Number

Filing Date

--	--	--

7 Inventorship

The applicant(s) are the sole inventors/joint inventors

Yes ☐

No ☒

8 Checklist

Claims 2 /

Abstract 1 /

Continuation sheets /

Description 16 /

Drawings 9 + 9

Priority Documents

Translations of Priority Documents

Patents Form 7/77

Patents Form 9/77

Patents Form 10/77

9 Request

We request the grant of a patent on the basis of this application

Signed: *Origin*
(Origin Limited)

Date: 24 April 99

Networked delivery of media files to clients**Technical Field**

5 This invention relates to methods for the delivery of media files, such as video and audio files, from a server to one or more clients using a network, in which the files stored on the server are encoded so that the generation and/or reception of their associated bitstreams can be controlled to yield a file at a client which meets the specific criteria (for example, image quality) of that client

10 Background to the Invention

One of the problems of delivering media over networks is that the encodings traditionally used to represent media take no account of the bandwidth limitations of networks, or of the varied and possibly conflicting requirements made upon the media by the shared access of multiple clients. Most media file formats have an "all-or-nothing" approach to the way they
15 are viewed: until the media is completely delivered and fully decoded the user cannot decide whether it is what they wanted. If not, then there is no option but to discard it completely. This is obviously an inefficient use of resources.

Recent work has tried to improve the usability of media over networks, and reduce inefficient use of network resources, by introducing "progressive" download. One example
20 is found in the Encoder Scalability Option in the Indeo product from Intel Corporation.

In such schemes the media is encoded as an ordered set of discrete blocks. Each block encodes the entire media in such a way that reception of the first block allows some aspect of the entire media object to be viewed, and each successive block received produces a discrete step in the quality of the representation at the client. The client can exert some
25 control over the use of network resource, and over the quality of the media delivered, by stopping the delivery when the representation has the desired quality.

The client, however, does not have control over the content of the block and therefore over the characteristics of the progressive improvement. Such decisions are made by the content provider when the media file is compiled. Neither does the client have any control over the
30 ordering of the blocks or over the ordering of data within the blocks.

Thus a strictly sequential approach both to content delivery and to the quality progression is enforced, and the parameters which control these aspects are not available to the client.

Also, the quality improvements available from the system are only available in fixed steps, the number of such steps also being predetermined by the content provider.

Another approach to meeting this need for 'progressive image transmission' is disclosed in US Patent No. 5880856 to Microsoft Corporation, which teaches a particular approach to the use of wavelet transforms. In the US 5880856 patent, an image is transformed using the wavelet transformation to yield 4 or 5 decomposition levels, with a base decomposition level giving a low resolution image, and increasingly higher decomposition levels giving higher resolution images. The client receives initially only the base decomposition level data, but the low resolution image resulting from the base decomposition level data gradually sharpens up as higher decomposition levels are received and decoded. Sharpening up of the image occurs as a result of 2 factors: first, as all of the 4 sub-bands which form each decomposition level are received and decoded, the image quality level increases slightly. As successive decomposition levels are received and decoded, the image quality increases more significantly. One characteristic of the system taught in the 5880856 patent is that the sub-bands and decomposition levels are all transmitted to clients in the same, fixed order, which is relatively inflexible.

This scheme allows the client full control over quality but provides no other parameters that can be controlled.

New applications for media delivery over networks, however, require much more flexibility and greater interactivity than is possible with any of the schemes described. Such an application is that of professional media browsing where the aim is rapidly to find, examine and annotate many different aspects or views of the media. Such a system must be capable of the following:-

- (a) Randomly to access any component of the media, where a component can be a time-ordered sequence of media parts, a media part sampled at a particular time, a spatial area within a media part, a colour within a media part, or any other aspect which can be described.
- (b) To access the above components at a quality, spatial resolution or frequency resolution specified by the client
- (c) To allow the results of any such access to be refined in such a way as to increase the information content and consequently improve the visual quality, amount of detail, or other aspect, of the material

- (d) To deliver such initial data or refinement data without redundancy, i.e., no information is sent which is not needed to fulfil the client requirements

Further reference may be made to the paper Beong-Jo Kim, Zixiang Xiong and William Pearlman, "*Very Low Bit-Rate Embedded Coding with 3D Set Partitioning in Hierarchical*

- 5 *Trees*", submitted to IEEE Trans. Circuits & Systems for Video Technology, Special Issue on Image and Video Processing for Emerging Interactive Multimedia Services, Sept. 1998. This paper discloses applying a SPIHT compression scheme to a wavelet-based transform, which yields a bitstream encoding multiple spatial resolutions, with progressive quality
10 is the use of 'flags' inserted by the client decoder into the received bitstream to mark temporal/spatial locations defined by the input resolution parameters. But because the 'flags' are inserted at the decoder, much (and in some cases, all) of the bitstream has to be received and stored at the client, so that considerable bandwidth may be wasted.

- 15 In order to satisfy all the requirements (a) to (d) a system design is required which integrates the design of the media encoder, bitstream, server and client.

Summary of the Invention

- 20 In accordance with the present invention, there is provided a method of encoding a media bitstream as a source file which can be reconstructed as a media file at a device, in which the source file comprises several layers, each layer carrying information which can be used to initialise or refine a particular aspect of the media file and where layer signalling information which identifies individual layers is appended to the layers as encoding proceeds.

- 25 In this specification, a 'layer' is a structure with a particular format, for example a hierarchical, block of data required to initialise or refine an encoded representation of media (for example an image, an image sequence, or an audio waveform). Hence, it may describe a media object in terms of further layers, each of which describes a part of a media object. These concepts will be discussed in greater detail in the following Detailed Description section.

- 30 The present invention may be implemented on a client server network, in which the source file resides on a server which is connected to one or more clients over a network. The term network should be expansively construed to cover any kind of data connection between 2 or more devices. A file is any consistent set of data, so that a media file is a consistent set of

data representing one or more media samples, such as frames of video or audio levels. The media file may be a sequential, fully embedded file.

In another aspect, there is provided a method of compiling a multi-layered source file into several sequential embedded representations with properties which satisfy criteria specified by one or more clients, where the sequential embedded representation allows initialisation of, or refinement to, any aspect of the reconstructed media.

Preferably, the sequential embedded representation is included in information existing at a client using a simple append operation for corresponding layers, in order to obtain refinement of a reconstructed media file at the client.

The sequential embedded representations may contain no information that does not directly improve the representation of a media file in a manner required by the client.

Preferably, the embedded representation also has the property whereby it can be included in information existing at the client for the same layer using a simple append operation, in order to obtain refinement of the reconstructed media. Additionally, it may contain no information that does not directly improve the representation in the manner required, and therefore allows efficient use of network resources.

In a preferred embodiment, the dynamic provision of progressive media file download for several clients with different requirements is achieved using a source file created using an encoder only once; there is no need to re-encode the media bitstream to satisfy differing criteria from several clients, as is needed in prior art systems.

The present invention is therefore predicated on the insight that:

(1) An embedded bitstream produces optimal image reconstruction for a given bit-rate, on sequential access and termination of that sequential access at an arbitrary point (the term 'embedding' is defined in the Detailed Description, Appendix 1);

(2) A layered bitstream offers flexibility as regards accessing information relevant to a particular client, but does not provide the benefits of (1), because while the layers themselves contain embedded data, the layers are not and cannot be sequentially ordered in an way which satisfies the needs of multiple clients in a distributed environment.

(3) That the insertion of layer signalling information at the encoder to identify individual layers enables the layered bitstream to be compiled into a fully embedded bitstream with the particular properties required by the client, prior to delivery to the client across a shared communications medium.

Such an embedded bitstream has particular benefits for the delivery of refinement information to a client in the context of unstructured, or non-linear access to a media file, where a human operator typically is interacting with a browse client. These benefits become manifest in the circumstances where a browse client tries to use the system resources in an optimal way by predicting how best to improve the media currently stored at the client by using any spare capacity in the communications channel. The human operator, however, can always defeat such predictions (for example by deciding to examine a totally different media file). In this case any transfer of refinement information can be abruptly terminated and the information already delivered can be added to that already stored, and used if required without causing errors or artefacts.

Brief Description of the Drawings

The invention will be described with reference to the accompanying drawings, in which:

- 15 Figure 1 is a schematic representing the sub-bands which result from applying wavelet transforms to an image in a conventional multi-scale decomposition;
- Figure 2 is a schematic representing the sub-bands which can be re-constructed in a multi-scale reconstruction according to the present invention;
- Figure 3 is a schematic representation of a network used in performing the method of media file delivery according to the present invention;
- 20 Figure 4 is a schematic representation of an encoder used in performing the method of media file delivery according to the present invention;
- Figure 5 is a schematic representation of a media server used in performing the method of media file delivery according to the present invention;
- 25 Figure 6 is a schematic representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers;
- Figure 7 is a schematic representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers in which all the layer information has been transmitted and used exactly to reconstitute the original
- 30 Figures 8 and 9 are schematics representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers in which only selected scales have been used,

resulting in an image with reduced spatial resolution and frequency content (a reduced-scale image).

Figure 10 is a schematic representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers in which information from all scales has been transmitted so that full spatial resolution is obtained, but the significance layers have been truncated so the image is distorted because lower-energy wavelet coefficients have not been transmitted to their full precision.

Figure 11 is a schematic representing a hierarchical decomposition of a particular media object (a video clip) into a set of layers in which a truncated significance layer has been transmitted for all scales, except those which encode horizontal details, which accordingly appear in the reconstructed image to the original fidelity.

Detailed Description.

Some key concepts

Wavelets and SPIHT Compression

The wavelet transform has been intensely studied for many years. Reference may for example be made to Mallat, Stephane G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, No.7, pp 674-692 (Jul 1989). The function of the wavelet transform is to compact most of the energy in an image to a small number of pixels. The transform also produces a set of sub-bands consisting mainly of zero or low-value coefficients; it therefore generates a large number of pixels with small or zero values. Because of the large number of pixels with small or zero values, there is considerable scope for a compression.

SPIHT compression is under increasing scrutiny: see for example Beong-Jo Kim, Zixiang Xiong and William Pearlman, "Very Low Bit-Rate Embedded Coding with 3D Set Partitioning in Hierarchical Trees", submitted to IEEE Trans. Circuits & Systems for Video Technology, Special Issue on Image and Video Processing for Emerging Interactive Multimedia Services, Sept. 1998.

The SPIHT compression algorithm is particularly effective operating in conjunction with a wavelet transform, since it can readily exploit the wavelet transform output to get high levels

of data reduction. Other compression schemes can also be used in the present invention, as will be appreciated by the skilled implementer.

The preferred wavelet transform in the present invention is Mallat's Fast Wavelet Transform (FWT), which generates a hierarchy of power-of-two images: it "critically" samples the

5 image at each stage of the transform. Every time that it generates a 2^n image, the spatial sampling frequency - the 'fineness' of detail which is represented - is reduced by a factor of two in x and y. To do this, the wavelet filter removes all signal energy above (sample freq/2), and distributes it into the high-frequency sub-bands. These are the 'detail' images which compress very well because they are mostly zeroes or low-values. The smaller,
10 critically-sampled (or smoothed) image which is generated at each stage can be viewed directly at that resolution because the high-frequencies which would otherwise cause aliasing have been removed. (Aliasing occurs when you try to render an image which contains frequency components too high for the display device. For example consider an image of a wooden fence where the uprights happen to lie on every other pixel. If you
15 simply reduce the resolution by two by missing out every other pixel the fence is rendered either as solid black or it disappears.)

Scale

FWT employs the concept of '*scale*'. A function can be described by analysing it into a set of discrete spatial/frequency components of differing *scales*.

20 In the preferred embodiment of the present invention, FWT is combined with SPIHT such that the image scale ordering resulting from the transform is maintained through the property of significance ordering of SPIHT.

The operation of FWT on an image can be seen in Figure 1. Here, the original of an image is transformed or decomposed into 4 sub-bands at Level 0. Each sub-band is a quarter the
25 size of the Original Image and a quarter the spatial resolution.

FWT operates as follows on a 2-D image:

- First, a high and low pass wavelet filter (i.e. a wavelet filter and a scaling function) are applied to each vertical line of pixels in the original, generating 2 sub-bands, one with high frequency information and the other with low frequency information.
- 30 • The filters are then applied to each horizontal line of pixels in the 2 sub-bands, splitting each sub-bands into two, yielding a total of 4 sub-bands, LL, LH, HL and HH. This is shown as Scale Level 0. LL contains low frequency information from the vertical and

horizontal convolutions. If viewed, it would look like the Original, but at a quarter resolution and a quarter size. LH contains low frequency information from the horizontal and high frequency from the vertical. HL contains high frequency from the horizontal and low frequency from the vertical. HH contains high frequency from the vertical and horizontal convolutions.

- A further decomposition can be performed on LL to yield a Scale Level 1 decomposition. In practice 4 to 7 successive decompositions may be applied.

Multi-Scale Reconstruction

Each subband describes the original image at a particular scale, with a particular sampling frequency. To reconstruct an image, inverse wavelet transforms are applied to the 4 sub-bands shown at Scale level 1 to re-generate sub-band LL at Level 0. Combining re-generated LL with the other 3 sub-bands in Level 0 enables the Original image to be obtained. As will be explained later, in the present invention, a *Layer Multiplexing* process is free to combine any sub-bands to produce a final image most appropriate to the particular requirements of the client. *Layer Truncation* can also optionally be used to limit the bit-rate at the cost of increased distortion.

Three examples are shown in Figure 2 which show the flexibility of the *Layer Multiplexing* approach. In the first example, Figure 2A, the LL sub-band of Level 1 is used: this will reconstruct a coarse version of the original, at one sixteenth the resolution and size of the original. In the second illustration, Figure 2B, the LL sub-band of Level 0 is used: this will reconstruct a larger and better quality image; it is one quarter the size and resolution of the original. In the third example, Figure 2C, the LL and LH sub-bands of Level 1 are used, together with the LH sub-band of Level 0. This reconstructs an original resolution image, but with horizontal features at all scales emphasised

System description

The system capable of performing the method of the present invention comprises a media encoder, network, media server and N clients, as shown in Figure 3. A media encoder processes the incoming media into a source file with a layered structure. The layering within the embedded bitstream may include any layering needed to support the differing download parameters which a client might specify to the server, such as *Region Layering* and *Significance/Scale Layering*. The encoder also inserts *Layer Signalling* prior to the

source file being transmitted as a bitstream over a digital communications network, where it is stored on a media server.

In Figure 3, there are N clients, each of which engages in a browse session with the media server during which media content and control information are transmitted. For each client a control channel to the server is maintained which allows that client to request that media be transmitted to it. The requests that can be made include (but are not limited to) the following:-

- Seek to a particular point in the media file.
- Transmit a media clip from the seek point at a specified bit-rate using the *Client Layer Template* to determine exactly what parts of the media are to be sent.
- Transmit a *Refinement Layer* for a previously sent media clip.
- Update or configure the *Client Layer Template* for this client.

These requests support a variety of functions in the client. For example, in a browsing application, a *Base Layer* may initially be sent to the client. This allows rapid scanning of low-quality material using minimum system resources (channel capacity, decoder CPU, etc.). To play a section with greater quality, the client updates the *Client Layer Template* to specify the exact refinement information which should now be delivered to improve the media information already stored at the client. The embedded nature of the data means that no processing is required at the client other than simply to insert the new data into the correct positions in the stored bitstream. This refinement information may perform (but is not limited to) the following actions:-

- Increase the temporal resolution of the media (i.e., for video, allow more individual frames to be resolved).
- Increase the spatial resolution of the video.
- Increase the sampling frequency of the audio.
- Decrease the degree of distortion of the media.
- Improve the quality of a 2-dimensional area within the image.
- Improve the media content at particular scales (for example enhance fine horizontal details in the video, or emphasise high frequencies in audio).

All these actions can be used to support application-level tasks such as video previewing using fast-forward and rewind, single-frame stepping forwards and backwards, and freeze-frame.

5 Description of the Encoder

As depicted in Figure 4, a media capture engine samples and stores incoming media at the source resolution. This is passed to a wavelet transform engine which performs *Multi-Scale Decomposition* of the source material into a wavelet coefficient representation.

A SPIHT encoder compresses the wavelet coefficients into an embedded bitstream. The bitstream encodes the original picture to a fidelity as near lossless as possible, subject to processing and network constraints. A *Layer Signalling* engine inserts information which delimits the individual blocks of data in the significance/scale layered bitstream. The signalling allows any subband and any refinement level to be efficiently accessed by a *Layer Multiplexing* process on the media server.

Description of Layer Signalling

Each Layer is labelled with a number of pieces of information to allow the server to deliver the data according to the client requirements. The information may include, but is not limited to, the following:

- (a) The size of the Layer
- (b) The type of Layer
- (c) A unique identifier or sequence number for the Layer data.
- (d) If the Layer is a spatial or temporal Scale Layer, as defined below, a value indicating the spatial or temporal resolution of the original media object from which this scale Layer is derived.
- (e) If the Layer is a spatial or temporal Scale Layer derived from a wavelet transform, as defined below, a value indicating the level and subband for this scale.
- (f) If the Layer is a spatial Region Layer, as defined below, a value indicating the size, shape and position of the region.

- (g) If the Layer is a temporal Region Layer, as defined below, a value indicating the frame or range of frames included in the region.
- (h) If the layer is a colour component Layer, as defined below, a value indicating the particular component.

5

Description of Client Layer Template

When a client requests the delivery of data it specifies a Client Layer Template that may include any of the following:

- (a) Parameters that specify which Layers are to be compiled into the bitstream.
- 10 (b) Parameters that specify the rate at which the data is to be delivered
- (c) Parameters that specify the progressive order in which the data is to be transmitted
- (d) Parameters that specify the set of data that is already held by the client and therefore does not need to be transmitted.
- (e) The maximum amount of data to be transmitted.

15

Description of the Media Server

The embedded bitstream is stored as a file on the Media Server.

An indexing process scans the bitstream as it is input to produce a cross-reference file which relates layer data to positions in the file. This is used by the Layer Multiplexer efficiently to create bitstreams with different properties for different clients.

20

When a client requests media data the Layer Multiplexer first of all consults the Client Layer Template for that individual client to discover which layers of data are required, and at what bit-rate. It then consults the Index Data to retrieve file pointers for the required layers. The requested data is then transmitted to the client over the digital communications network.

25

In detail, this process involves the following steps, in which the Server:

- (a) locates the Client Layer Template for the client,
- (b) determines the total set of Layers to be transmitted,
- (c) determines the order in which the Layers are to be transmitted,

- (d) locates Layer data which may be in fast temporary storage, such as cache memory, as a result of delivery to another client,
- (e) locates the remaining Layer data using file pointers derived from the index data,
- (f) delivers the Layer data in order, at the specified rate,
- 5 (g) terminates the transmission when the specified maximum amount of data has been reached.

Further Layer Multiplexing examples

Figure 6 shows an example of a hierarchical decomposition of a particular media object (a video clip) into a set of layers. The video is first of all transformed into one or more temporal regions (T-regions), where each T-region can encode different numbers of frames, with different encoding parameters, and different underlying layer structures. The ability to generate T-regions with different properties allows the encoder to respond to external stimuli and to adapt to new situations. For example in a surveillance application a signal from a motion detector can start the generation of a new T-region with a greater amount of information and resolving power in the underlying layers.

T-regions may be composed of one or more temporal scale (T-scale) layers, where a Scale Layer is a layer for which all the refinement information is for a particular Scale.

Thus each T-scale captures a particular aspect of the temporal activity in the T-region.

20 Each T-scale may be composed of one or more spatial regions (S-regions) for which all the underlying information is for a particular 2-dimensional area in the image plane. This enables different areas to be encoded with differing information content; in particular it allows 'hi-fidelity' windows to be defined over areas of particular interest.

Each S-region may be composed of one or more spatial scale (S-scale) layers, where each S-scale captures a particular aspect of the spatial activity in the S-region. Each S-scale may be composed of one or more component layers, where all the underlying information refers to a particular component in a colour system, for example, R,G,B or Y,U,V or other format.

Finally, the component layer is composed of one or more significance layers, as follows:

30 In an embedded, bit-plane oriented compression scheme such as SPIHT, compression is obtained by partially ordering all the samples with respect to bit-significance, i.e., the position at which the highest-order bit in a sample is set. Since this corresponds to the

magnitude of the sample, the samples are effectively ordered with respect to their energies, or the contribution they make to the reconstructed object. A significance layer is generated by choosing a bit-position and outputting the value of that bit-position (1 or 0) for all the samples for which that bit-position is defined. Hence, in Figure 6, 'R' is the most significant bit of a component's coefficients and 'r' is the bit-significance at which the encoder stops refinement.

An embedded representation is thus obtained by starting with the most-significant bits of the coefficients, and sending successively lower-order bits, until a desired quality is reached. From the point of view of a particular sample, the sample is first approximated by having its most-significant bit set, and successively approximated to greater precision as lower-significance layers become available. The process of increasing the precision of a sample is called refinement.

An encoding of source material can be structured in different ways, and use different combinations of layers, while delivering exactly the same information. For example, a representation of an entire video at low resolution may come before other layers, each of which causes a step in the quality of the entire video. Alternatively all the information for a particular frame may appear before any information for the next in which case reading the representation sequentially will produce one frame after the other, each delivered to as high a quality as possible. However, given signalling to delimit individual layers all structures are equivalent since one can be processed into another.

The primary use of layer signalling is to allow a hierarchical, structured and layered representation to be compiled into a plurality of sequential, unstructured and embedded representations for use by multiple clients. Figures 7 to 11 are examples of such representations which may be obtained from a single layered object. In figure 7 all the layer information has been transmitted and used exactly to reconstitute the original. In figures 8 and 9 only selected scales have been used and result in an image with reduced spatial resolution and frequency content (a reduced-scale image). In figure 10 information from all scales has been transmitted so full spatial resolution is obtained, but the significance layers have been truncated so the image is distorted because lower-energy wavelet coefficients have not been transmitted to their full precision. In figure 11 a truncated significance layer has been transmitted for all scales, except those which encode horizontal details, which accordingly appear in the reconstructed image to the original fidelity.

Appendix 1

Some Definitions of Terms, as used in the context of the illustrated embodiments

5 *Scale*

A function can be analysed into a set of components with different time/space/frequency content. These components are called scales, and the process of analysis into scales is called multi-scale decomposition. The analysis is performed by a waveform of limited duration and zero integral, called a wavelet. Components that are highly localised in space or time but have an ill-defined frequency spectrum are small-scale and capture fine detail. Components that are spread through space but have a precisely-defined frequency spectrum are large-scale and capture general trends.

Embedding

15 Encoding a source object in such a way that the generation and/or reception of the encoded bitstream can be terminated at any point, the resulting representation being the 'best' attainable, according to some criteria set by the encoder.

Layer

20 A layer is structure with a particular format: a media stream may be organised by an encoder into a tree-structured hierarchy of Layers, where a Layer represents a node in the tree and its associated subtree. The subtree consists, either of further Layers, or of a Simple Layer which consists of an embedded bitstream.

25 *Base Layer*

A layer which is sent first to a client to initialise a representation of a media object at that client, to which refinement layers are added. The Base Layer usually carries 'coarse-scale' or 'trend' information which allows a general impression of the media to be obtained, although any layer can be used as a Base Layer.

Refinement Layer

A Layer which is sent to a client subsequent to a Base Layer and which reduces the distortion of the representation of a media object at that client. The Refinement Layer usually carries 'fine-scale' or 'detail' information which enhances the representation existing at the client.

Simple Layer

A layer which contains embedded refinement information only, i.e., there are no further sub-layers.

Layer Embedding

The process of writing Layers to a bitstream in such a way that those layers carrying the 'best' information, according to some criteria which may be set by the encoder, appear early in the sequence.

Significance Layer

A layer where all the refinement information refers to a particular bit-position for all the coefficients undergoing refinement.

Scale Layer

A layer for which all the refinement information is for a particular scale.

Region Layer

A layer for which all the refinement information is for a particular connected region in a space or time-varying function. This enables different regions to be studied in different ways, with information being concentrated in different combinations of Layers according to the requirements of the bitstream. For example, a 'hi-fidelity' window can be defined over areas of particular interest.

Significance-Scale Layering.

A layering scheme whereby Significance Layers are embedded within Scale Layers. Such a scheme is important because it allows a trade-off to be made between bit-rate, spatial resolution and distortion in the bitstream transmitted to the client.

5

Distortion

The distortion of an image can be measured in terms of its Peak Signal-to-Noise Ratio (PSNR) where $PSNR = 10\log(255^2/MSE)$ dB, and MSE is the image's Mean Squared Error.

10

Claims

- 5 1. A method of encoding a media bitstream as a source file which can be reconstructed as a media file at a device, in which the source file comprises several layers, each layer carrying information which can be used to initialise or refine a particular aspect of the media file and where layer signalling information which identifies individual layers is appended to the layers as encoding proceeds.
- 10 2. The method of Claim 1 wherein a particular aspect of the media file is temporal resolution, spatial resolution or image sample distortion.
3. The method of Claim 1 wherein the several layers include layers relevant to region layering, scale layering, significance layering or significance-scale layering.
4. The method of any preceding claim in which the encoded file is generated using a
15 wavelet transform and subsequent compression.
5. The method of Claim 4 in which the subsequent compression is SPIHT compression.
6. The method of Claim 1 in which any one or more of the variables of spatial resolution, quality and bit rate can be set at each device, and only those layers are sent to a
20 device which result in meeting the value of the variables specified by each client
7. The method of Claim 1 wherein any one or more of spatial resolution, quality and bit rate can be set at each device when initiating or during the playing of a media file or sequence of media files.
8. The method of Claim 1 wherein the layers represent several sub-bands and the
25 method includes the step of combining sub-bands to produce the media file received by the device, any sub-band being combinable.
9. The method of Claim 8 wherein several sub-bands are combined so that the media file received by the device satisfies criteria specified by that device.

10. The method of Claim 1 wherein a client has associated with it a template defining those layers which are required by it and that template can be accessed by a computer operable to generate a bitstream conforming to the template layer profile.

5 11. The method of Claim 10 wherein layer signalling information in a bitstream generated by an encoder are indexed and stored and that indexed information is used to generate a bitstream transmitted to a client.

10 12. A method of compiling a multi-layered source file into a sequential embedded representation with properties which satisfy criteria specified by any of a plurality of clients, where the sequential embedded representation allows initialisation of, or refinement to, any aspect of the reconstructed media.

13. The method of Claim 12 in which the sequential embedded representation is included in information existing at a client using a simple append operation for corresponding layers, in order to obtain refinement of a reconstructed media file at the client.

15 14. The method of Claim 13 in which the sequential embedded representation contains no information that does not directly improve the representation of a media file in a manner required by the client.

15. A media file which is deliverable using any of the above methods.

20 16. A computer program which when running on a client enables the client to receive and playback a media file delivered using any of the above methods.

17. A computer program which when running on a server or encoder enables the server or encoder to perform any of the above methods.

Networked delivery of media files to clients

Abstract

5 The delivery of a single encoded video or audio file which can be played back over a
network to various clients at different data rates, resolutions and quality levels, as
individually determined by each client, is disclosed. An encoder inserts into the media file
bitstreams 'layer signalling' information which delimits and identifies a number of different
layers (for example, different 'significance/scale layers' and different 'region layers'). A
media server stores the media files, and can distribute to different networked clients
10 bitstreams with different properties depending upon the layers which are requested by or are
appropriate to each client.



BLANK PAGE

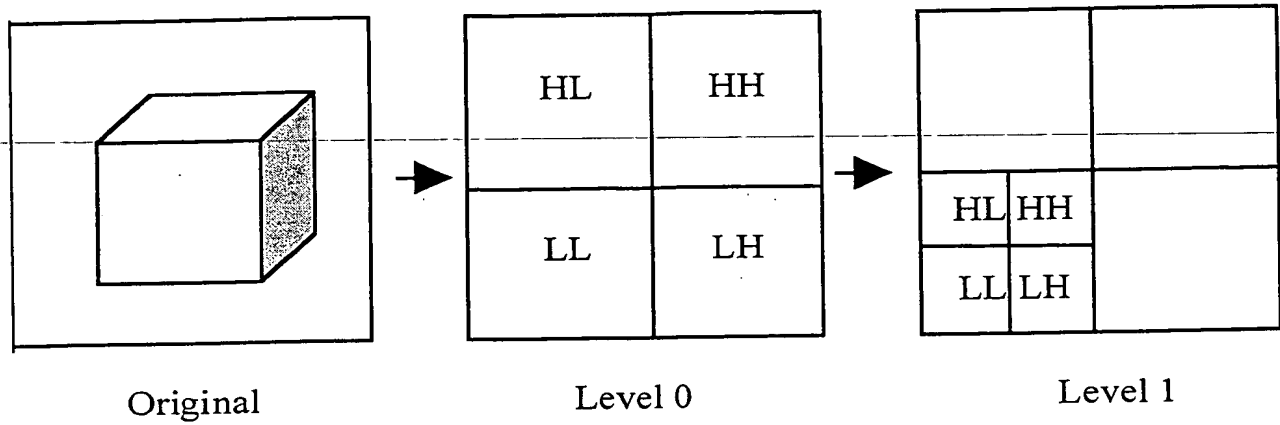


Figure 1

Figure 2

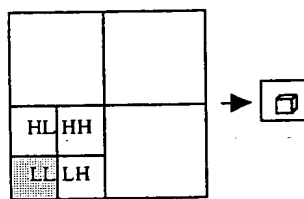


Fig 2A

Reconstruction
using LL
Subband of level 1
(lowest-resolution)

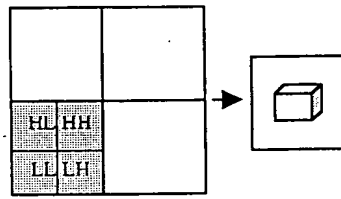


Fig.2B

Reconstruction
using LL
Subband of level 0
(medium
resolution)

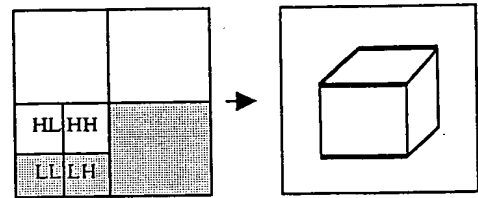


Fig 2C

Reconstruction using LL & LH,
subband
of level 1
And LH subband of level 0
(Original resolution,
horizontal features at all scales
emphasised)



BLANK PAGE

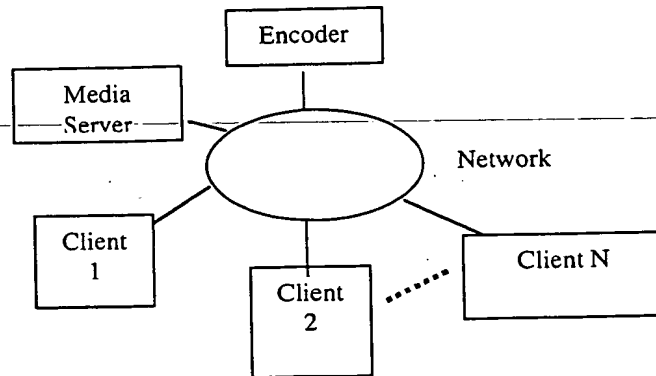


Figure 3

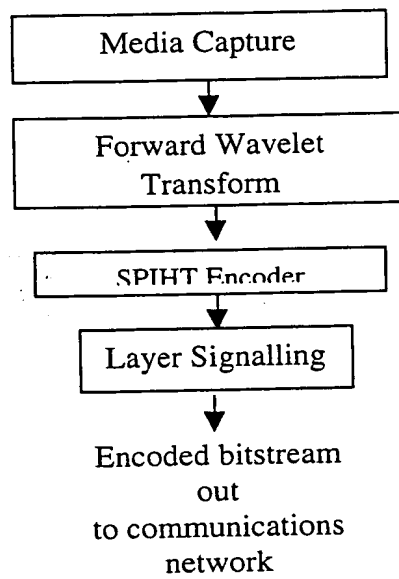
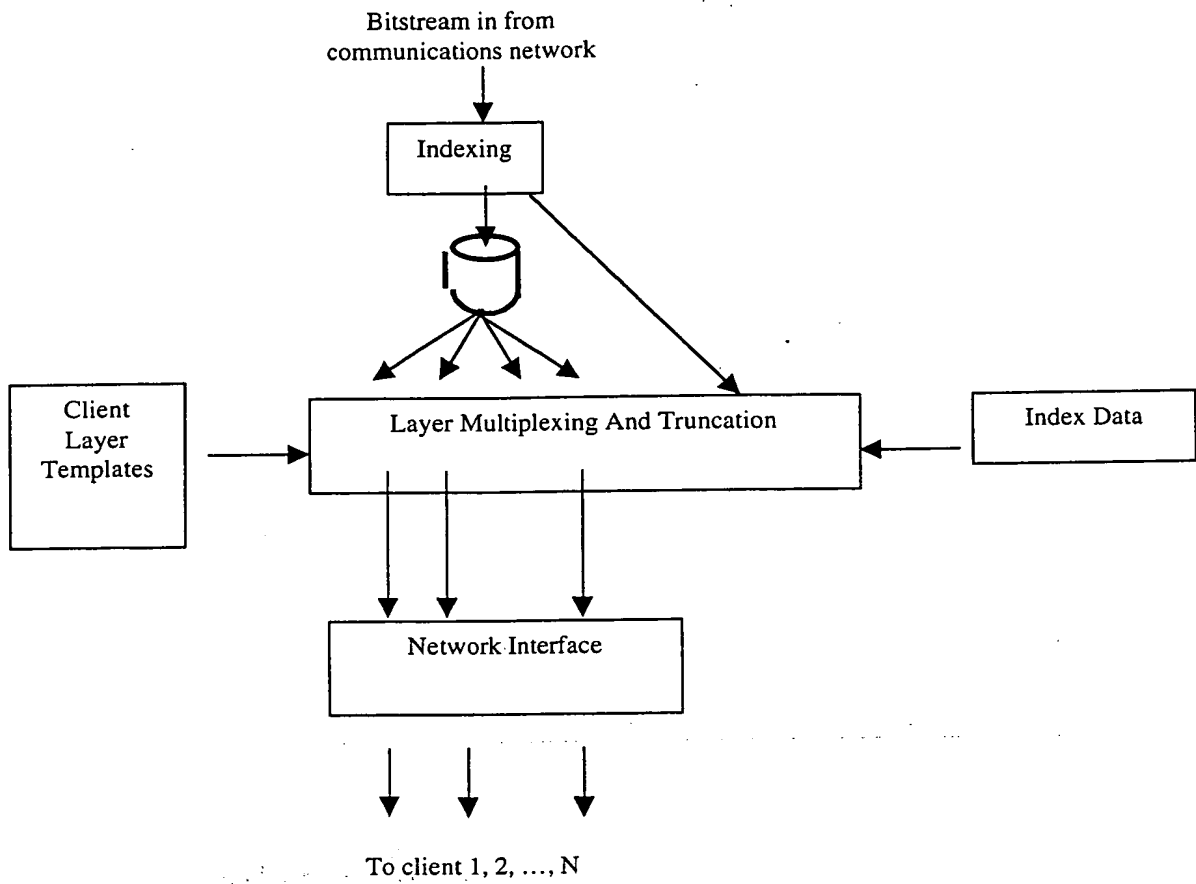


Figure 4



BLANK PAGE

Figure 5



BLANK PAGE

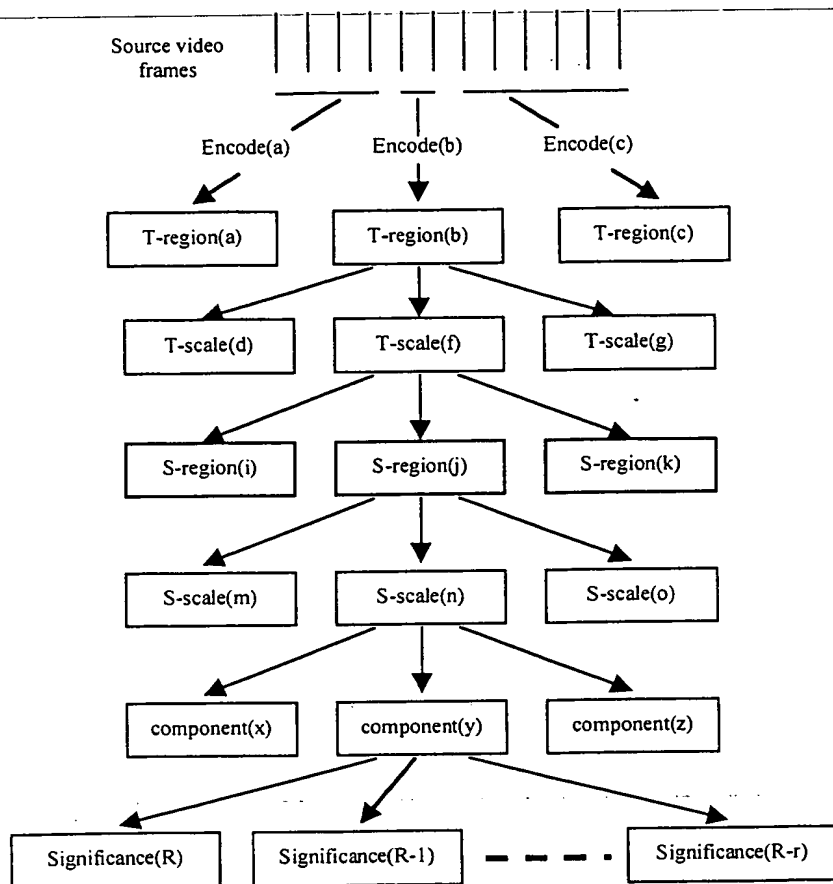


Figure 6



BLANK PAGE

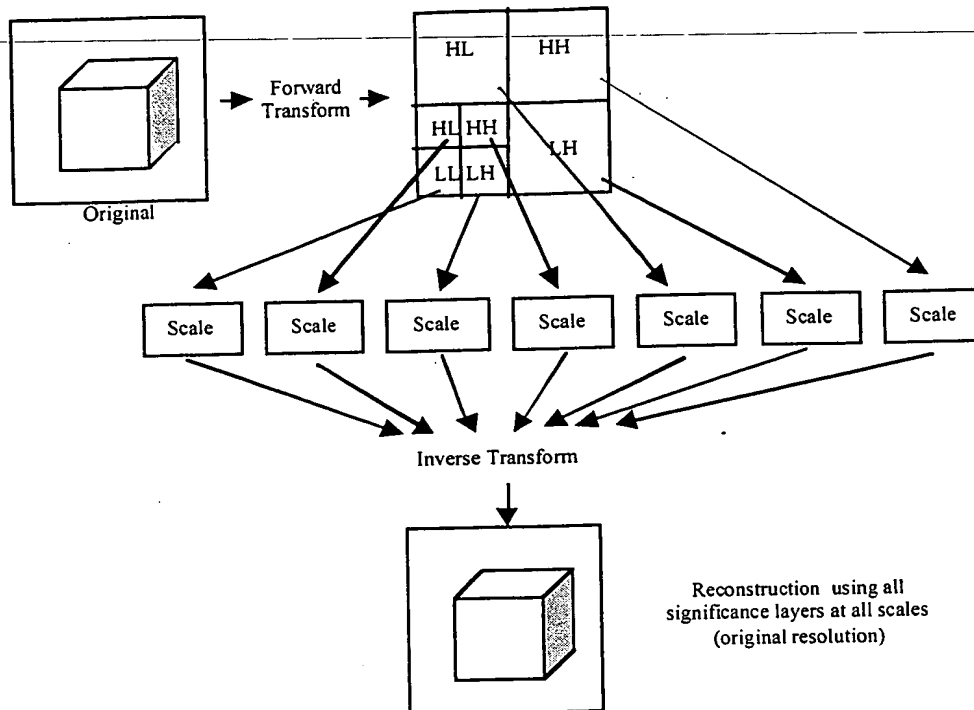


Figure 7



BLANK PAGE

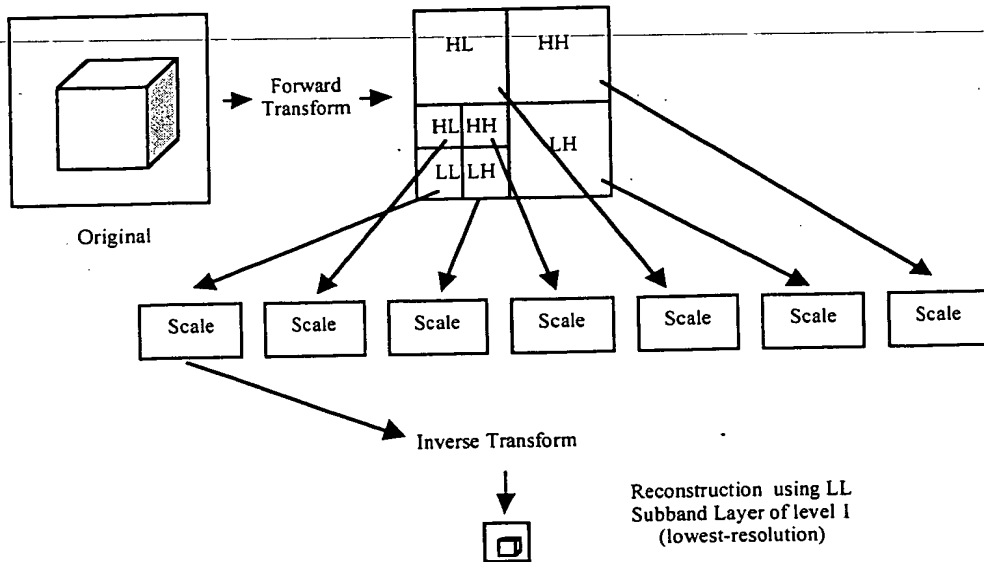


Figure 8



BLANK PAGE

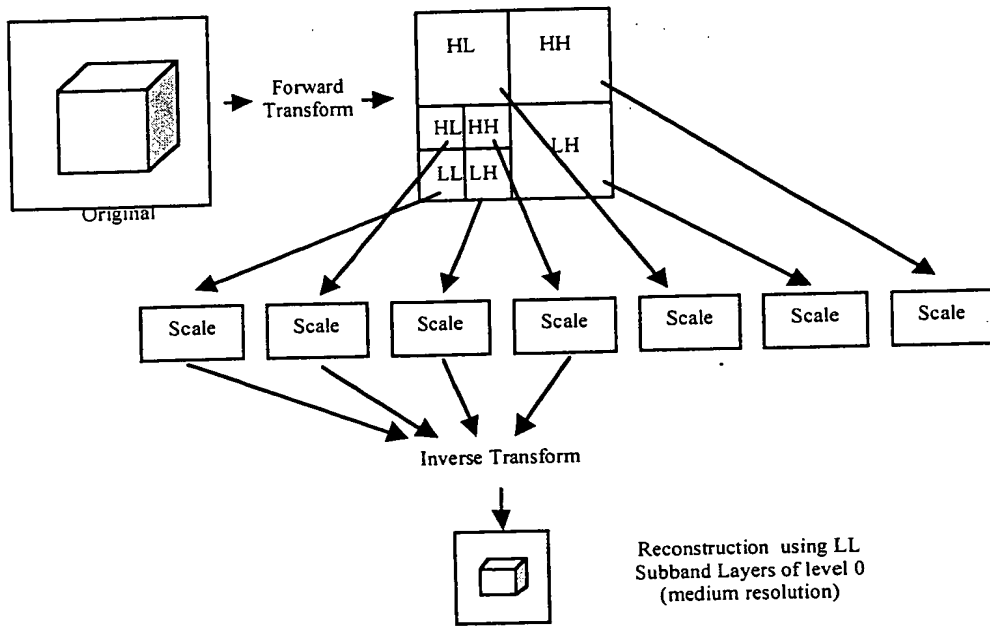


Figure 9

BLANK PAGE

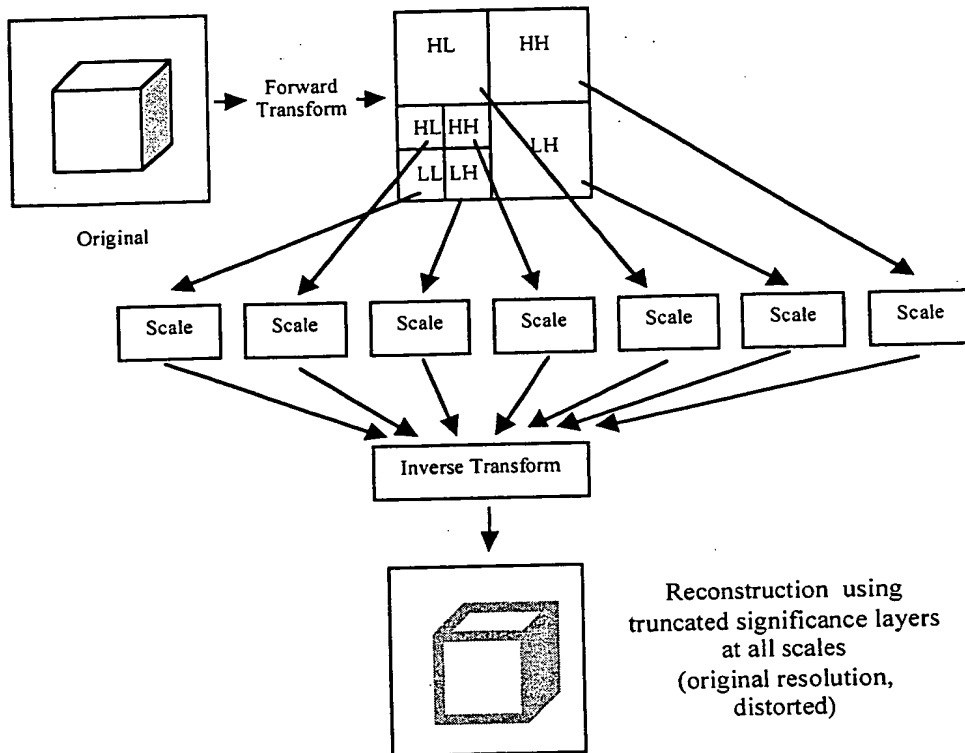


Figure 10

BLANK PAGE

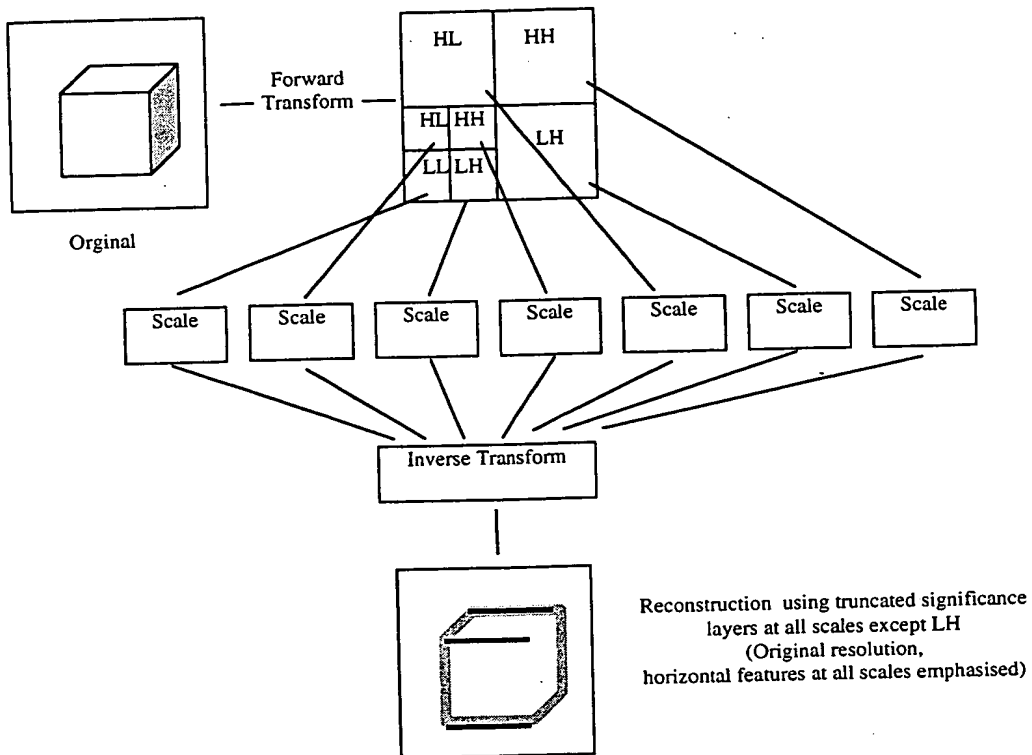


Figure 11

PC T'gB0010/614
ORIGIN

19-7-00

BLANK PAGE